Integrating Genomics, Proteomics, and Bioinformatics

John N. Weinstein, M.D., Ph.D.

As we enter what has been called the "Post-Genomic" Era in biomedical sciences (I would prefer "Omic Era") [1-3], new demands are being placed on bioinformatics. For most omic studies, there are two phases: Analysis and Interpretation [4]. The analysis phase is principally the province of statisticians and experts in machine learning; the interpretation phase is principally the province of biologists, preferably with encyclopedic knowledge of the tools and database resources necessary to put whole lists of genes and gene products into perspective.

Our own work along these lines has principally related to the 60 cell lines (the NCI-60) [5-7] used in the National Cancer Institute's drug discovery program. These cells have been treated with >70,000 compounds one at a time and independently over a 12-year period. Cell lines in culture do not fully reflect cells *in vivo*, of course, but, historically, most of our knowledge of molecular pharmacology and targets has come from cultured cells, not clinical material. We and our colleagues have assessed expression patterns in the NCI-60 using 2-D protein gel electrophoresis [8], high density "reverse-phase" protein arrays, cDNA microarrays [9,10], and oligonucleotide chips [11]. To find patterns in the data, we have then developed new data visualizations, including the familiar Clustered Image Map [12], and a tool (MedMiner) that streamlines literature searches on genes and drugs [13]. We and our collaborators next characterized the cells at the DNA level by comparative genomic hybridization (CGH), spectral karyotyping, array-CGH, and SNP chip and then developed algorithms and a program package called LeadScope/LeadMiner [14]. This package makes it possible to predict which molecular substructures will be found in drugs that are active against cells expressing large amounts of a selected gene – or vice versa. Finally, we have developed the program package GEEVS (GEnome Exploration and Visualization System) to integrate all of these disparate types of data at the DNA, RNA, protein, functional, and pharmacological levels. Part of the GEEVS package, and also for independent use, are MatchMiner [15] and GoMiner, two programs of major utility that facilitate biological interpretation of omic information. One clinically interesting pharmacogenomic [10] outcome of these various studies: L-asparaginase may prove useful for therapy of ovarian cancers that express only low levels of asparagine synthetase. See http://discover.nci.nih.gov.

[1] Weinstein, J.N., Fishing expeditions, Science 1998; 282: 628-629.

[2] Weinstein, J.N. 'Omic' and hypothesis-driven research in the molecular pharmacology of cancer. Current Opinion in Pharmacology 2002; 2:361-365.

[3] Weinstein, J.N. Pharmacogenomics: Teaching old drugs new tricks. New England J. Med. 2000; 343: 1408-1409. (Editorial).

[4] Weinstein, J.N., Scherf, U., Lee, J.K., Nishizuka, S., Gwadry, F., Ajay, Bussey, K., Kim, S., Smith, L.H., Tanabe,L., Richman, S., Alexander, J., Kouros-Mehr, H., Maunakea, A., and Reinhold, W.C. The bioinformatics of Microarray gene expression profiling. Cytometry 2002; 47:46-49.

[5] Boyd, M.R. and Paull, K.D. Some practical considerations and applications of the National Cancer Institute in vitro anticancer drug discovery screen. Drug Dev. Res. 34:91; 1995;

[6] Paull, K.D., et al., Display and analysis of patterns of differential activity of drugs against human tumor cell lines: development of mean graph and COMPARE algorithm. J. Natl. Cancer Inst. 81:1088, 1989;

[7] Weinstein, J.N., Kohn, K.W., Grever, M.R., Viswanadhan, V.N., Rubinstein, L.V., Monks, A., Scudiero, D.A., Welch, L., Koutsoukos, A.J., Chiausa, A.J., and Paull, K.D. Neural computing in cancer drug development: Predicting mechanism of action. Science, 1992; 258: 447-451.

[8] Myers, T.G., Waltham, M., Li, G., Buolamwini, J.K., Scudiero, D.A., Rubinstein, L.V., Paull, K.D., Sausville, E.A., and Anderson, N.L., and Weinstein, J.N. A protein expression database for the molecular pharmacology of cancer. Electrophoresis 1997; 18:647-653.

[9] Ross, D.T., Scherf, U., Eisen, M.B., Perou, C.M., Spellman, P., Iyer, V., Jeffrey, S.S., Van de Rijn, M., Waltham, M., Pergamenschikov, A., Lee, J.C.F., Lashkari, D., Shalon, D., Myers, T.G., Weinstein, J.N., Botstein, D., and Brown, P.O. Systematic variation in gene expression patterns in human cancer cell lines. Nature Genetics 2000; 24: 227-235.

[10] Scherf, U., Ross, D.T., Waltham, M., Smith, L.H., Lee, J.K., Kohn, K.W., Reinhold, W.C., Myers, T.G., Andrews, D.T., Scudiero, D.A., Eisen, M.B., Sausville, E.A., Pommier, Y., Botstein, D., Brown, P.O., and Weinstein, J.N. A cDNA microarray gene expression database for the molecular pharmacology of cancer. Nature Genetics 2000; 24: 236-244.

[11] Staunton, J.E., Slonim, D.K., Coller, H.A., Tamayo, P., Angelo, M.J., Park, J., Scherf, U., Lee, J.K., Weinstein, J.N., Mesirov, J.P., Lander, E.S., and Golub, T.R. Chemosensitivity prediction by transcriptional profiling. Proc. Natl. Acad. Sci. U.S.A. 2001; 98:10787-10792.

[12] Weinstein, J.N., Myers, T.G., O'Connor, P.M., Friend, S.H., Fornace, A.J., Kohn, K.W., Fojo, T., Bates, S.E., Rubinstein, L.V., Anderson, N.L., Buolamwini, J.K., van Osdol, W.W., Monks, A.P., Scudiero, D.A., Sausville, E.A., Zaharevitz, D.W., Bunow, B., Viswanadhan, V.N., Johnson, G.S., Wittes, R.E., and Paull, K.D. An information-intensive approach to the molecular pharmacology of cancer. Science 1997; 275:343-349.

[13] Tanabe, L., Smith, L.H., Lee, J.K., Scherf, U., Hunter, L., and Weinstein, J.N. MedMiner: An internet tool for filtering and organizing biomedical information, with application to gene expression profiling. BioTechniques. 1999; 27: 1210-1217.

[14] Blower, P., Jr., Yang, C., Fligner, M.A., Verducci, J.S., Yu, L., Richman, S., and Weinstein, J.N. Pharmacogenomic Analysis: Correlating molecular substructure classes with microarray gene expression data. The Pharmacogenomics Journal (Nature) 2002; 2:259-271.

[15] Bussey, K.J., Kane, D., Sunshine, M., Narasimhan, S., Nishizuka, S., Reinhold, W.C., Zeeberg, B., Ajay, Weinstein, J.N. MatchMiner: A tool for batch navigation among gene and gene product identifiers. Genome Research, submitted.